Laboratory 9
Data Structures Representation
Computer Science 240

Writing X86 Code Directly

Assembly Directives
Begin with a dot and indicate structural information useful to the assembler, linker, or debugger.

- indicates label *main* is a global symbol that can be accessed by other code modules.

```
.globl main
```

- store instructions following directive in the text segment of memory

```
.text
```

- store declared data following directive in the data segment of memory

```
.data
```

- allocate space for an 8 byte value and initialize to 0

```
i:   .quad 0
```

- allocate space and initialize to specified null-terminated string

```
str: .string  "Enter a value"
```

Reference: http://tigcc.ticalc.org/doc/gnuasm.html#SEC67

simple.c: (C code)

```
#include <stdio.h>


long total = 0;


I
int sum(int x,int y) {
  int t = x + y;
  total +=t;
  return t;
}

int main() {
   int x = 2;
   int y = 3;
   printf("Sum =
%d\n",sum(x,y));
   printf("Total =
%d\n",total);
   return 0;
}
```

simple.s: (X86 code)

```
          .data //use the data segment of
memory

total:   .quad   0 //8 bytes with initial value 0
fstr1:   .string  "Sum =  %d\n"
fstr2:    .string  "Total = %d\n"

          .text
          .globl main
sum:
          lea  (%rsi,%rdi,1),%eax
          add   %eax,total
          ret


main:
        mov    $0x3,%esi
        mov    $0x2,%edi
        call     sum

        mov    %eax,%esi
        mov    $fstr1,%edi
        mov    $0x0,%eax
        call    printf

        mov    $total,%esi
        mov    $fstr2,%edi
        mov    $0x0,%eax
        call    printf

        mov    $0x0,%eax
        ret
```

```c
#include <stdio.h>

int z;



int square(int n) {
  return n*n;
}

int main() {

   int x = square(3);


   int y = square(4);


   z = x + y;



 printf("Calculation produces %d\n",z);

 return 0;
}
```

```
        .data

z:      .long 0
fstr:   .string "Calculation produces
%d\n"

        .text
        .globl main

square: mov %edi,%eax
        imul %edi,%eax
        ret

main:   push %rbx
        mov $3,%edi
        call square
        mov %eax,%ebx
        mov $4,%edi
        call square
        lea (%ebx,%eax,1),%esi
        mov %esi,z
        mov $fstr,%edi
        mov $0,%eax
        call printf
        mov $0,%eax
        pop %rbx
        ret
```

One-dimensional arrays

Different languages use different implementations at the machine level to represent data structures.

In Java, arrays are actually implemented as arrays of addresses (pointers) to the elements, which are stored elsewhere in memory (not necessarily in contiguous locations).

In C, the elements of the array are stored in a contiguous block, starting at the base address of the array.

In the C model,

address of element in array = base address + element size * index

If the size of the element is limited to 1, 2, or 4 bytes, what is another more efficient way to accomplish the multiplication?

In C, to define some arrays of 8 elements of different sizes:

```
long  qelements[] = {0xF,  0xE,  0xD,  0xC};
int   elements[]  = {0x1,  0x3,  0x5,  0x7,  0x9,  0x11, 0x13, 0x15};
short welements[] = {0x23, 0x25, 0x27, 0x29, 0x31, 0x33, 0x35, 0x37}
char  belements[] = {0x20, 0x30, 0x40, 0x50, 0x60, 0x70, 0x80, 0x90}
```

The equivalent in X86 is:

```
     .data
qelements: .quad 0xF,0xE,0XD,0XC
elements:  .long 0x1, 0x3, 0x5, 0x7, 0x9, 0x11,0x13,0x15
welements: .word 0x23,0x25,0x27,0x29,0x31,0x33,0x35,0x37
belements: .byte 0x20,0x30,0x40,0x50,0x60,0x70,0x80,0x90
```

The contents of memory starting at *qelements* displayed using *gdb* would look something like this:

```
0x0049700 <qelements>:     0x00000000 0x0000000F 0x00000000 0x0000000E
0x8049714 <qelements+16>:  0x00000000 0x0000000D 0x00000000 0x0000000C
0x0049724 <elements >:     0x00000001 0x00000003 0x00000005 0x00000007
0x0049734 <elements+16>:   0x00000009 0x00000011 0x00000013 0x00000015
0x0049744 <welements  >:   0x00250023 0x00290027 0x00330031 0x00370035
0x 049754 <belements  >:   0x50403020 0x90807060
```

Two-dimensional arrays

In C, when nested array of arrays are used, each row is stored contiguously in memory (*row-major* format), and the address of an element can be calculated by the following formula (size of row is the number of columns in a row):

```
address of element[row][col] =
              base address of array +
              (row * size of row * size of element) +
              (col * size of element)

         -or-

              base address of array +
              (row*size of row + col)*size of element
```

In C, to define a 4x4 array of integers:

```
int twodarr[4][4] =  {{0x1, 0x2, 0x3, 0x4},
                      {0x4, 0x6, 0x7, 0x8},
                      {0x9, 0x10,0x11,0x12},
                      {0x13,0x14,0x15,0x16}};
```

The equivalent in X86 is:

```
           .data
twodarr:   .long 0x1, 0x2, 0x3, 0x4
           .long 0x5, 0x6, 0x7, 0x8
           .long 0x9, 0x10,0x11,0x12
           .long 0x13,0x14,0x15,0x16
```

Either would be displayed using *gdb* as:

```
0x80497a0 <twodarr   >:    0x00000001 0x00000002 0x00000003 0x00000004
0x80497b0 <twodarr+16>:    0x00000005 0x00000006 0x00000007 0x00000008
0x80497c0 <twodarr+32>:    0x00000009 0x00000010 0x00000011 0x00000012
0x80497d0 <twodarr+48>:    0x00000013 0x00000014 0x00000015 0x00000016
```

Dark Buffer Arts

```
/* functions in umbrella.c (given) used for Exploit 1 */

void smoke()
{
  entry_check(0);  /* Make sure entered this function properly */
  printf("Smoke!: You called smoke()\n");
  validate(0);
  exit(0);
}


void test()
{
  unsigned long long val;
  volatile unsigned long long local = 0xdeadbeef;
  char* variable_length;
  entry_check(3);  /* Make sure entered this function properly */
  val = getbuf();
  if (val <= 40) {
      variable_length = alloca(val);
  }
  entry_check(3);
  /* Check for corrupted stack */
  if (local != 0xdeadbeef) {
      printf("Sabotaged!: the stack has been corrupted\n");
  }
  else if (val == cookie) {
      printf("Boom!: getbuf returned 0x%llx\n", val);
      if (local != 0xdeadbeef) {
        printf("Sabotaged!: the stack has been corrupted\n");
      }
      validate(3);
  }
  else {
      printf("Dud: getbuf returned 0x%llx\n", val);
  }
}

unsigned long long getbuf()
{
  char buf[36];
  volatile char* variable_length;
  int i;
  unsigned long long val = (unsigned long long)Gets(buf);
  variable_length = alloca((val % 40) < 36 ? 36 : val % 40);
  for(i = 0; i < 36; i++)
  {
      variable_length[i] = buf[i];
  }
  return val % 40;
}
```

```
Dump of assembler code for function getbuf:
   0x0000000000400dd0 <+0>:   push   %rbp
   0x0000000000400dd1 <+1>:   mov    %rsp,%rbp
   0x0000000000400dd4 <+4>:   sub    $0x30,%rsp
   0x0000000000400dd8 <+8>:   lea    -0x30(%rbp),%rdi
   0x0000000000400ddc <+12>:  callq  0x400cb0 <Gets>
   0x0000000000400de1 <+17>:  movabs $0xcccccccccccccccd,%rdx
   0x0000000000400deb <+27>:  mov    %rax,%rcx
   0x0000000000400dee <+30>:  mul    %rdx
   0x0000000000400df1 <+33>:  mov    %rdx,%rax
   0x0000000000400df4 <+36>:  mov    $0x24,%edx
   0x0000000000400df9 <+41>:  shr    $0x5,%rax
   0x0000000000400dfd <+45>:  lea    (%rax,%rax,4),%rax
   0x0000000000400e01 <+49>:  shl    $0x3,%rax
   0x0000000000400e05 <+53>:  sub    %rax,%rcx
   0x0000000000400e08 <+56>:  cmp    $0x24,%rcx
   0x0000000000400e0c <+60>:  mov    %rcx,%rax
   0x0000000000400e0f <+63>:  cmovae %rcx,%rdx
   0x0000000000400e13 <+67>:  lea    -0x30(%rbp),%rcx
   0x0000000000400e17 <+71>:  add    $0x1e,%rdx
   0x0000000000400e1b <+75>:  and    $0xfffffffffffffff0,%rdx
   0x0000000000400e1f <+79>:  lea    0x24(%rcx),%r8
   0x0000000000400e23 <+83>:  sub    %rdx,%rsp
   0x0000000000400e26 <+86>:  lea    0xf(%rsp),%rsi
   0x0000000000400e2b <+91>:  and    $0xfffffffffffffff0,%rsi
   0x0000000000400e2f <+95>:  nop
   0x0000000000400e30 <+96>:  movzbl (%rcx),%edi
   0x0000000000400e33 <+99>:  add    $0x1,%rcx
   0x0000000000400e37 <+103>: add    $0x1,%rsi
   0x0000000000400e3b <+107>: mov    %dil,-0x1(%rsi)
   0x0000000000400e3f <+111>: cmp    %r8,%rcx
   0x0000000000400e42 <+114>: jne    0x400e30 <getbuf+96>
   0x0000000000400e44 <+116>: leaveq
   0x0000000000400e45 <+117>: retq
End of assembler dump.
```

Gets

The C library function **char *gets(char *str)** reads a line from stdin and stores it into the string pointed to by str. It stops when either the newline character is read or when the end-of-file is reached, whichever comes first.

Declaration

Following is the declaration for gets() function.

```
char *gets(char *str)
```

Parameters

- **str** − This is the pointer to an array of chars where the C string is stored.

Return Value

This function returns str on success, and NULL on error or when end of file occurs, while no characters have been read.

Example

The following example shows the usage of gets() function.

```
#include <stdio.h>

int main () {
   char str[50];

   printf("Enter a string : ");
   gets(str);

   printf("You entered: %s", str);

   return(0);
}
```

In Exploit 1, your task is to get `umbrella` to execute the code for `smoke()` when `getbuf()` executes its return statement, rather than returning to `test()`.

You can do this by supplying an exploit string that overwrites the stored return pointer in the stack frame for `getbuf()` with the address of the first instruction in `smoke`.

Note that your exploit string may also corrupt other parts of the stack state, but this will not cause a problem, because `smoke()` causes the program to exit directly.

1.Determine **how many** bytes and what **hexadecimal value** each byte should be. Create a text file using Emacs called *exploit.hex* containing a string which represents your bytes.

- Each byte should be a two-digit value
- Avoid using '0A' as a value
- Avoid entering the Return key. If your string is longer that one line, just let the line wrap.

EXAMPLE: if your exploit needs to be 5 bytes long, and the last byte has to be a 0x42, then your file would contain:

```
01 02 03 04 05 06 07 08 09 42
```

In this example, only the last byte needs to be a specific value, so it is useful to give the other bytes values in numeric order (it makes it easy to verify that you have 9 bytes before the 42).

The following is a **poor** choice of values for the non-specific bytes, because you can easily make a mistake in your count:

```
00 00 00 00 00 00 00 00 00 42
```

2. Convert the file into hexadecimal bytes by using the *hex2raw* tool:

```
$ ./hex2raw < exploit1.hex > exploit1.bytes
```

3. Run the program with your exploit:

```
$ ./umbrella -u your_cs_username < exploit1.bytes
```